CHAPTER 18

18.1. a. This follows from equation (18.5). First, $\mathrm{E}(\bar{y}_1) = \mathrm{E}(y|w = 1)$ and $\mathrm{E}(\bar{y}_0) = \mathrm{E}(y|w = 1)$. Therefore, by (18.5),

$$\mathrm{E}(\bar{y}_1 - \bar{y}_0) = [\mathrm{E}(y_0|w = 1) - \mathrm{E}(y_0|w = 0)] + ATE_1,$$

and so the bias is given by the first term.

b. If $\mathrm{E}(y_0|w = 1) < \mathrm{E}(y_0|w = 0)$, those who participate in the program would have had lower average earnings without training than those who chose not to participate. This is a form of sample selection, and, on average, leads to an underestimate of the impact of the program.

18.3. The following Stata session estimates α using the three different regression approaches. It would have made sense to add *unem74* and *unem75* to the vector $\mathbf{x}$, but I did not do so:

```
. probit train re74 re75 age agesq nodegree married black hisp

Iteration 0:   log likelihood =     -302.1
Iteration 1:   log likelihood = -294.07642
Iteration 2:   log likelihood = -294.06748
Iteration 3:   log likelihood = -294.06748

Probit estimates                                  Number of obs   =        445
                                                  LR chi2(8)      =      16.07
                                                  Prob > chi2     =     0.0415
Log likelihood = -294.06748                       Pseudo R2       =     0.0266

------------------------------------------------------------------------------
       train |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        re74 |  -.0189577   .0159392    -1.19   0.234    -.0501979    .0122825
        re75 |   .0371871   .0271086     1.37   0.170    -.0159447     .090319
         age |  -.0005467   .0534045    -0.01   0.992    -.1052176    .1041242
       agesq |   .0000719   .0008734     0.08   0.934    -.0016399    .0017837
    nodegree |    -.44195   .1515457    -2.92   0.004    -.7389742   -.1449258
     married |    .091519   .1726192     0.53   0.596    -.2468083    .4298464
       black |  -.1446253   .2271609    -0.64   0.524    -.5898524    .3006019
```

```
        hisp |  -.5004545   .3079227    -1.63   0.104    -1.103972    .1030629
       _cons |   .2284561   .8154273     0.28   0.779    -1.369752    1.826664
------------------------------------------------------------------------------

. predict phat
(option p assumed; Pr(train))

. sum phat

    Variable |     Obs        Mean   Std. Dev.       Min        Max
-------------+-----------------------------------------------------
        phat |     445    .4155321    .0934459   .1638736   .6738951

. gen traphat0 = train*(phat - .416)

. reg unem78 train phat

      Source |       SS       df       MS              Number of obs =     445
-------------+------------------------------           F(  2,   442) =    3.13
       Model |   1.3226496     2  .661324802           Prob > F      =  0.0449
    Residual |  93.4998223   442   .21153806           R-squared     =  0.0139
-------------+------------------------------           Adj R-squared =  0.0095
       Total |  94.8224719   444  .213564126           Root MSE      =  .45993

------------------------------------------------------------------------------
      unem78 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       train |   -.110242    .045039    -2.45   0.015    -.1987593   -.0217247
        phat |  -.0101531   .2378099    -0.04   0.966    -.4775317    .4572254
       _cons |   .3579151   .0994803     3.60   0.000     .1624018    .5534283
------------------------------------------------------------------------------

. reg unem78 train phat traphat0

      Source |       SS       df       MS              Number of obs =     445
-------------+------------------------------           F(  3,   441) =    2.84
       Model |  1.79802041     3  .599340137           Prob > F      =  0.0375
    Residual |  93.0244515   441  .210939799           R-squared     =  0.0190
-------------+------------------------------           Adj R-squared =  0.0123
       Total |  94.8224719   444  .213564126           Root MSE      =  .45928

------------------------------------------------------------------------------
      unem78 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       train |  -.1066934   .0450374    -2.37   0.018     -.195208   -.0181789
        phat |   .3009852   .3151992     0.95   0.340    -.3184939    .9204644
    traphat0 |   -.719599   .4793509    -1.50   0.134    -1.661695     .222497
       _cons |    .233225    .129489     1.80   0.072    -.0212673    .4877173
------------------------------------------------------------------------------

. reg unem78 train re74 re75 age agesq nodegree married black hisp
```

```
      Source |       SS       df       MS              Number of obs =     445
-------------+------------------------------           F(  9,   435) =    2.75
       Model |  5.09784844     9  .566427604           Prob > F      =  0.0040
    Residual |  89.7246235   435  .206263502           R-squared     =  0.0538
-------------+------------------------------           Adj R-squared =  0.0342
       Total |  94.8224719   444  .213564126           Root MSE      =  .45416

------------------------------------------------------------------------------
      unem78 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       train |  -.1105582   .0444832    -2.49   0.013    -.1979868   -.0231295
        re74 |  -.0025525   .0053889    -0.47   0.636    -.0131441    .0080391
        re75 |   -.007121   .0094371    -0.75   0.451     -.025669    .0114269
         age |   .0304127   .0189565     1.60   0.109    -.0068449    .0676704
       agesq |  -.0004949   .0003098    -1.60   0.111    -.0011038    .0001139
    nodegree |   .0421444   .0550176     0.77   0.444    -.0659889    .1502777
     married |  -.0296401   .0620734    -0.48   0.633    -.1516412    .0923609
       black |    .180637   .0815002     2.22   0.027     .0204538    .3408202
        hisp |  -.0392887   .1078464    -0.36   0.716    -.2512535    .1726761
       _cons |  -.2342579   .2905718    -0.81   0.421    -.8053572    .3368413
------------------------------------------------------------------------------
```

In all three cases, the average treatment effect is estimated to be right around -.11: participating in job training is estimated to reduce the unemployment probability by about .11. Of course, in this example, training status was randomly assigned, so we are not surprised that different methods lead to roughly the same estimate. An alternative, of course, is to use a probit model for *unem78* on *train* and **x**.

18.5. a. I used the following Stata session to answer all parts:

```
. probit train re74 re75 age agesq nodegree married black hisp

Iteration 0:   log likelihood =     -302.1
Iteration 1:   log likelihood = -294.07642
Iteration 2:   log likelihood = -294.06748
Iteration 3:   log likelihood = -294.06748

Probit estimates                                  Number of obs   =        445
                                                  LR chi2(8)      =      16.07
                                                  Prob > chi2     =     0.0415
Log likelihood = -294.06748                       Pseudo R2       =     0.0266

------------------------------------------------------------------------------
```

```
       train |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        re74 |  -.0189577   .0159392    -1.19   0.234    -.0501979    .0122825
        re75 |   .0371871   .0271086     1.37   0.170    -.0159447     .090319
         age |  -.0005467   .0534045    -0.01   0.992    -.1052176    .1041242
       agesq |   .0000719   .0008734     0.08   0.934    -.0016399    .0017837
    nodegree |    -.44195   .1515457    -2.92   0.004    -.7389742   -.1449258
     married |    .091519   .1726192     0.53   0.596    -.2468083    .4298464
       black |  -.1446253   .2271609    -0.64   0.524    -.5898524    .3006019
        hisp |  -.5004545   .3079227    -1.63   0.104    -1.103972    .1030629
       _cons |   .2284561   .8154273     0.28   0.779    -1.369752    1.826664
------------------------------------------------------------------------------

. predict phat
(option p assumed; Pr(train))


. reg re78 train re74 re75 age agesq nodegree married black hisp (phat re74
re75 age agesq nodegree married black hisp)

Instrumental variables (2SLS) regression

      Source |       SS       df       MS              Number of obs =     445
-------------+------------------------------           F(  9,   435) =    1.75
       Model |  703.776258     9  78.197362            Prob > F      =  0.0763
    Residual |  18821.8804   435  43.2686905           R-squared     =  0.0360
-------------+------------------------------           Adj R-squared =  0.0161
       Total |  19525.6566   444  43.9767041           Root MSE      =  6.5779

------------------------------------------------------------------------------
        re78 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       train |   .0699177   18.00172     0.00   0.997    -35.31125    35.45109
        re74 |   .0624611   .1453799     0.43   0.668    -.2232733    .3481955
        re75 |   .0863775   .2814839     0.31   0.759    -.4668602    .6396151
         age |   .1998802   .2746971     0.73   0.467    -.3400184    .7397788
       agesq |  -.0024826   .0045238    -0.55   0.583    -.0113738    .0064086
    nodegree |  -1.367622   3.203039    -0.43   0.670    -7.662979    4.927734
     married |   -.050672   1.098774    -0.05   0.963    -2.210237    2.108893
       black |  -2.203087   1.554259    -1.42   0.157    -5.257878    .8517046
        hisp |  -.2953534   3.656719    -0.08   0.936    -7.482387     6.89168
       _cons |   4.613857   11.47144     0.40   0.688    -17.93248     27.1602
------------------------------------------------------------------------------

. reg phat re74 re75 age agesq nodegree married black hisp

      Source |       SS       df       MS              Number of obs =     445
-------------+------------------------------           F(  8,   436) =69767.44
       Model |  3.87404126     8  .484255158           Prob > F      =  0.0000
    Residual |  .003026272   436  6.9410e-06           R-squared     =  0.9992
-------------+------------------------------           Adj R-squared =  0.9992
       Total |  3.87706754   444  .008732134           Root MSE      =  .00263
```

phat	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
re74	-.0069301	.0000312	-222.04	0.000	-.0069914	-.0068687
re75	.0139209	.0000546	254.82	0.000	.0138135	.0140283
age	-.0003207	.00011	-2.92	0.004	-.0005368	-.0001046
agesq	.0000293	1.80e-06	16.31	0.000	.0000258	.0000328
nodegree	-.1726018	.000316	-546.14	0.000	-.1732229	-.1719806
married	.0352802	.00036	98.01	0.000	.0345727	.0359877
black	-.0562315	.0004726	-118.99	0.000	-.0571603	-.0553027
hisp	-.1838453	.0006238	-294.71	0.000	-.1850713	-.1826192
_cons	.5907578	.0016786	351.93	0.000	.5874586	.594057

b. The IV estimate of α is very small -- .070, much smaller than when we used either linear regression or the propensity score in a regression in Example 18.2. (When we do not instrument for *train*, $\hat{\alpha}$ = 1.625, se = .640.) The very large standard error (18.00) suggests severe collinearity among the instruments.

c. The collinearity suspected in part b is confirmed by regressing $\hat{\Phi}_i$ on the $\mathbf{x}_i$: the *R*-squared is .9992, which means there is virtually no separate variation in $\hat{\Phi}_i$ that cannot be explained by $\mathbf{x}_i$.

d. This example illustrates why trying to achieve identification off of a nonlinearity can be fraught with problems. Generally, it is not a good idea.

18.7. To be added.

18.9. a. We can start with equation (18.66),

$$y = \eta_0 + \mathbf{x}\boldsymbol{\gamma} + \beta w + w\cdot(\mathbf{x} - \boldsymbol{\psi})\boldsymbol{\delta} + u + w\cdot v + e,$$

and, again, we will replace $w\cdot v$ with its expectation given $(\mathbf{x},\mathbf{z})$ and an error. But $\mathrm{E}(w\cdot v|\mathbf{x},\mathbf{z}) = \mathrm{E}[\mathrm{E}(w\cdot v|\mathbf{x},\mathbf{z},v)|\mathbf{x},\mathbf{z}] = \mathrm{E}[\mathrm{E}(w|\mathbf{x},\mathbf{z},v)\cdot v|\mathbf{x},\mathbf{z}] = \mathrm{E}[\exp(\pi_0 + \mathbf{x}\boldsymbol{\pi}_1 + \mathbf{z}\boldsymbol{\pi}_2 + \pi_3 v)\cdot v|\mathbf{x},\mathbf{z}] = \xi\cdot\exp(\pi_0 + \mathbf{x}\boldsymbol{\pi}_1 + \mathbf{z}\boldsymbol{\pi}_2)$ where $\xi = \mathrm{E}[\exp(\pi_3 v)\cdot v]$, and we have used the assumption that v is independent of $(\mathbf{x},\mathbf{z})$. Now, define $r = u + [w - \mathrm{E}(w\cdot v|\mathbf{x},\mathbf{z})] + e$. Given the assumptions, $\mathrm{E}(r|\mathbf{x},\mathbf{z}) = 0$. [Note that we do not need to replace $\boldsymbol{\pi}_0$ with a different constant, as is implied in the statement of the problem.] So we can write

$$y = \eta_0 + \mathbf{x}\boldsymbol{\gamma} + \beta w + w\cdot(\mathbf{x} - \boldsymbol{\psi})\boldsymbol{\delta} + \xi\mathrm{E}(w|\mathbf{x},\mathbf{z}) + r,\ \mathrm{E}(r|\mathbf{x},\mathbf{z}) = 0.$$

b. The ATE β is not identified by the IV estimator applied to the extended equation. If $h \equiv h(\mathbf{x},\mathbf{z})$ is any function of $(\mathbf{x},\mathbf{z})$, $\mathrm{L}(w|1,\mathbf{x},q,h) = \mathrm{L}(w|q) = q$ because $q = \mathrm{E}(w|\mathbf{x},\mathbf{z})$. In effect, becaue we need to include $\mathrm{E}(w|\mathbf{x},\mathbf{z})$ in the estimating equation, no other functions of $(\mathbf{x},\mathbf{z})$ are valid as instruments. This is a clear weakness of the approach.

c. This is not what I intended to ask. What I should have said is, assume we can write $w = \exp(\boldsymbol{\pi}_0 + \mathbf{x}\boldsymbol{\pi}_1 + \mathbf{z}\boldsymbol{\pi}_2 + g)$, where $\mathrm{E}(u|g,\mathbf{x},\mathbf{z}) = \boldsymbol{\rho}\cdot g$ and $\mathrm{E}(v|g,\mathbf{x},\mathbf{z}) = \theta\cdot g$. These are standard linearity assumptions under independence of (u,v,g) and $(\mathbf{x},\mathbf{z})$. Then we take the expected value of (18.66) conditional on $(g,\mathbf{x},\mathbf{z})$:

$$\begin{aligned}\mathrm{E}(y|v,\mathbf{x},\mathbf{z}) &= \eta_0 + \mathbf{x}\boldsymbol{\gamma} + \beta w + w\cdot(\mathbf{x} - \boldsymbol{\psi})\boldsymbol{\delta} + \mathrm{E}(u|g,\mathbf{x},\mathbf{z}) + w\mathrm{E}(v|g\mathbf{x},\mathbf{z}) \\ &\quad + \mathrm{E}(e|g,\mathbf{x},\mathbf{z}) \\ &= \eta_0 + \mathbf{x}\boldsymbol{\gamma} + \beta w + w\cdot(\mathbf{x} - \boldsymbol{\psi})\boldsymbol{\delta} + \boldsymbol{\rho}\cdot g + \theta w\cdot g,\end{aligned}$$

where we have used the fact that w is a function of $(g,\mathbf{x},\mathbf{z})$ and $\mathrm{E}(e|g,\mathbf{x},\mathbf{z}) = 0$. The last equation suggests a two-step procedure. First, since $\log(w_i) = \boldsymbol{\pi}_0 + \mathbf{x}_i\boldsymbol{\pi}_1 + \mathbf{z}_i\boldsymbol{\pi}_2 + g_i$, we can consistently estimate $\boldsymbol{\pi}_0$, $\boldsymbol{\pi}_1$, and $\boldsymbol{\pi}_2$ from the OLS regression $\log(w_i)$ on 1, $\mathbf{x}_i$, $\mathbf{z}_i$, $i = 1,\ldots,N$. From this regression, we need the residuals, $\hat{g}_i$, $i = 1,\ldots,N$. In the second step, run the regression

$$y_i \text{ on } 1,\ \mathbf{x}_i,\ w_i,\ w_i(\mathbf{x}_i - \bar{\mathbf{x}}),\ \hat{g}_i,\ w_i\hat{g}_i,\ i = 1,\ldots,N.$$

As usual, the coefficient on w_i is the consistent estimator of β, the average treatment effect. A standard joint significant test -- for example, an F-type test -- on the last two terms effectively tests the null hypothesis that w is exogenous.

CHAPTER 19

19.1. a. This is a simple problem in univariate calculus. Write $q(\mu) \equiv \mu_o \log(\mu) - \mu$ for $\mu > 0$. Then $dq(\mu)/d\mu = \mu_o/\mu - 1$, so $\mu = \mu_o$ uniquely sets the derivative to zero. The second derivative of $q(\mu)$ is $-\mu_o\mu^{-2} > 0$ for all $\mu > 0$, so the sufficient second order condition is satisfied.

b. For the exponential case, $q(\mu) \equiv E[\ell_i(\mu)] = -\mu_o/\mu - \log(\mu)$. The first order condition is $\mu_o\mu^{-2} - \mu^{-1} = 0$, which is uniquely solved by $\mu = \mu_o$. The second derivative is $-2\mu_o\mu^{-3} + \mu^{-2}$, which, when evaluated at μ_o, gives $-2\mu_o^{-2} + \mu_o^{-2} = -\mu_o^{-2} < 0$.

19.3. The following is Stata output used to answer parts a through f. The answers are given below.

```
. reg cigs lcigpric lincome restaurn white educ age agesq

      Source |       SS       df       MS              Number of obs =     807
-------------+------------------------------           F(  7,   799) =    6.38
       Model |  8029.43631     7  1147.06233           Prob > F      =  0.0000
    Residual |  143724.246   799  179.880158           R-squared     =  0.0529
-------------+------------------------------           Adj R-squared =  0.0446
       Total |  151753.683   806  188.280003           Root MSE      =  13.412

------------------------------------------------------------------------------
        cigs |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lcigpric |  -.8509044   5.782321    -0.15   0.883    -12.20124    10.49943
     lincome |   .8690144   .7287636     1.19   0.233     -.561503    2.299532
    restaurn |  -2.865621   1.117406    -2.56   0.011    -5.059019   -.6722235
       white |  -.5592363   1.459461    -0.38   0.702    -3.424067    2.305594
        educ |  -.5017533   .1671677    -3.00   0.003     -.829893   -.1736136
         age |   .7745021   .1605158     4.83   0.000     .4594197    1.089585
```

```
       agesq |  -.0090686   .0017481    -5.19   0.000    -.0124999   -.0056373
       _cons |  -2.682435   24.22073    -0.11   0.912    -50.22621    44.86134
------------------------------------------------------------------------------

. test lcigpric lincome

 ( 1)  lcigpric = 0.0
 ( 2)  lincome = 0.0

       F(  2,   799) =    0.71
            Prob > F =    0.4899


. reg cigs lcigpric lincome restaurn white educ age agesq, robust

Regression with robust standard errors                 Number of obs =     807
                                                       F(  7,   799) =    9.38
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0529
                                                       Root MSE      =  13.412

------------------------------------------------------------------------------
             |               Robust
        cigs |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lcigpric |  -.8509044   6.054396    -0.14   0.888     -12.7353     11.0335
     lincome |   .8690144    .597972     1.45   0.147    -.3047671    2.042796
    restaurn |  -2.865621   1.017275    -2.82   0.005    -4.862469   -.8687741
       white |  -.5592363   1.378283    -0.41   0.685     -3.26472    2.146247
        educ |  -.5017533   .1624097    -3.09   0.002    -.8205533   -.1829532
         age |   .7745021   .1380317     5.61   0.000     .5035545     1.04545
       agesq |  -.0090686   .0014589    -6.22   0.000    -.0119324   -.0062048
       _cons |  -2.682435   25.90194    -0.10   0.918    -53.52632    48.16145
------------------------------------------------------------------------------

. test lcigpric lincome

 ( 1)  lcigpric = 0.0
 ( 2)  lincome = 0.0

       F(  2,   799) =    1.07
            Prob > F =    0.3441


. poisson cigs lcigpric lincome restaurn white educ age agesq

Iteration 0:   log likelihood = -8111.8346
Iteration 1:   log likelihood = -8111.5191
Iteration 2:   log likelihood =  -8111.519

Poisson regression                                Number of obs   =        807
                                                  LR chi2(7)      =    1068.70
```

```
                                                  Prob > chi2     =     0.0000
Log likelihood =  -8111.519                       Pseudo R2       =     0.0618

------------------------------------------------------------------------------
        cigs |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lcigpric |  -.1059607   .1433932    -0.74   0.460    -.3870061    .1750847
     lincome |   .1037275   .0202811     5.11   0.000     .0639772    .1434779
    restaurn |  -.3636059   .0312231   -11.65   0.000    -.4248021   -.3024098
       white |  -.0552012   .0374207    -1.48   0.140    -.1285444    .0181421
        educ |  -.0594225   .0042564   -13.96   0.000    -.0677648   -.0510802
         age |   .1142571   .0049694    22.99   0.000     .1045172    .1239969
       agesq |  -.0013708    .000057   -24.07   0.000    -.0014825   -.0012592
       _cons |   .3964494   .6139626     0.65   0.518    -.8068952    1.599794
------------------------------------------------------------------------------

. glm cigs lcigpric lincome restaurn white educ age agesq, family(poisson)
sca(x2)

Iteration 0:   log likelihood = -8380.1083
Iteration 1:   log likelihood = -8111.6454
Iteration 2:   log likelihood =  -8111.519
Iteration 3:   log likelihood =  -8111.519

Generalized linear models                          No. of obs      =       807
Optimization     : ML: Newton-Raphson              Residual df     =       799
                                                   Scale param     =         1
Deviance         =  14752.46933                    (1/df) Deviance =  18.46367
Pearson          =  16232.70987                    (1/df) Pearson  =  20.31628

Variance function: V(u) = u                        [Poisson]
Link function    : g(u) = ln(u)                    [Log]
Standard errors  : OIM

Log likelihood   = -8111.519022                    AIC             =  20.12272
BIC              =  14698.92274

------------------------------------------------------------------------------
        cigs |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lcigpric |  -.1059607   .6463244    -0.16   0.870    -1.372733    1.160812
     lincome |   .1037275   .0914144     1.13   0.257    -.0754414    .2828965
    restaurn |  -.3636059   .1407338    -2.58   0.010    -.6394391   -.0877728
       white |  -.0552011   .1686685    -0.33   0.743    -.3857854    .2753831
        educ |  -.0594225   .0191849    -3.10   0.002    -.0970243   -.0218208
         age |   .1142571   .0223989     5.10   0.000     .0703561     .158158
       agesq |  -.0013708   .0002567    -5.34   0.000     -.001874   -.0008677
       _cons |   .3964493    2.76735     0.14   0.886    -5.027457    5.820355
------------------------------------------------------------------------------
(Standard errors scaled using square root of Pearson X2-based dispersion)

* The estimate of sigma is
```

```
. di sqrt(20.32)
4.5077711

. poisson cigs restaurn white educ age agesq

Iteration 0:   log likelihood =  -8125.618
Iteration 1:   log likelihood = -8125.2907
Iteration 2:   log likelihood = -8125.2906

Poisson regression                                Number of obs   =        807
                                                  LR chi2(5)      =    1041.16
                                                  Prob > chi2     =     0.0000
Log likelihood = -8125.2906                       Pseudo R2       =     0.0602

------------------------------------------------------------------------------
        cigs |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    restaurn |  -.3545336   .0308796   -11.48   0.000    -.4150564   -.2940107
       white |  -.0618025    .037371    -1.65   0.098    -.1350483    .0114433
        educ |  -.0532166   .0040652   -13.09   0.000    -.0611842   -.0452489
         age |   .1211174   .0048175    25.14   0.000     .1116754    .1305594
       agesq |  -.0014458   .0000553   -26.14   0.000    -.0015543   -.0013374
       _cons |   .7617484   .1095991     6.95   0.000     .5469381    .9765587
------------------------------------------------------------------------------

. di 2*(8125.291 - 8111.519)
27.544

. * This is the usual LR statistic.  The GLM version is obtained by
. * dividing by 20.32:

. di 2*(8125.291 - 8111.519)/(20.32)
1.3555118

. glm cigs lcigpric lincome restaurn white educ age agesq, family(poisson)
robust

Iteration 0:   log likelihood = -8380.1083
Iteration 1:   log likelihood = -8111.6454
Iteration 2:   log likelihood =  -8111.519
Iteration 3:   log likelihood =  -8111.519

Generalized linear models                          No. of obs      =        807
Optimization     : ML: Newton-Raphson              Residual df     =        799
                                                   Scale param     =          1
Deviance         =  14752.46933                    (1/df) Deviance =   18.46367
Pearson          =  16232.70987                    (1/df) Pearson  =   20.31628

Variance function: V(u) = u                        [Poisson]
Link function    : g(u) = ln(u)                    [Log]
Standard errors  : Sandwich
```

```
Log likelihood   = -8111.519022                   AIC              =  20.12272
BIC              =  14698.92274

------------------------------------------------------------------------------
             |               Robust
        cigs |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lcigpric |  -.1059607   .6681827    -0.16   0.874    -1.415575    1.203653
     lincome |   .1037275    .083299     1.25   0.213    -.0595355    .2669906
    restaurn |  -.3636059    .140366    -2.59   0.010    -.6387182   -.0884937
       white |  -.0552011   .1632959    -0.34   0.735    -.3752553     .264853
        educ |  -.0594225   .0192058    -3.09   0.002    -.0970653   -.0217798
         age |   .1142571   .0212322     5.38   0.000     .0726427    .1558715
       agesq |  -.0013708   .0002446    -5.60   0.000    -.0018503   -.0008914
       _cons |   .3964493    2.97704     0.13   0.894    -5.438442     6.23134
------------------------------------------------------------------------------

. di .1143/(2*.00137)
41.715328
```

a. Neither the price nor income variable is significant at any reasonable significance level, although the coefficient estimates are the expected sign. It does not matter whether we use the usual or robust standard errors. The two variables are jointly insignificant, too, using the usual and heteroskedasticity-robust tests (p-values = .490, .344, respectively).

b. While the price variable is still very insignificant (p-value = .46), the income variable, based on the usual Poisson standard errors, is very significant: t = 5.11. Both estimates are elasticities: the estimate price elasticity is −.106 and the estimated income elasticity is .104. Incidentally, if you drop *restaurn* -- a binary indicator for restaurant smoking restrictions at the state level -- then log(*cigpric*) becomes much more significant (but using the incorrect standard errors). In this data set, both *cigpric* and *restaurn* vary only at the state level, and, not surprisingly, they are significantly correlated. (States that have restaurant smoking restrictions also have higher average prices, on the order of 2.9%.)

c. The GLM estimate of σ is $\hat{\sigma} = 4.51$. This means all of the Poisson standard errors should be multiplied by this factor, as is done using the "glm" command in Stata, with the option "sca(x2)." The *t* statistic on *lcigpric* is now very small (-.16), and that on *lincome* falls to 1.13 -- much more in line with the linear model *t* statistic (1.19 with the usual standard errors). Clearly, using the maximum likelihood standard errors is very misleading in this example. With the GLM standard errors, the restaurant restriction variable, education, and the age variables are still significant. (Interestingly, there is no race effect, conditional on the other covariates.)

d. The usual *LR* statistic is 2(8125.291 - 8111.519) = 27.54, which is a very large value in a χ_2^2 distribution (*p*-value ≈ 0). The *QLR* statistic divides the usual *LR* statistic by $\hat{\sigma}^2 = 20.32$, so *QLR* = 1.36 (*p*-value ≈ .51). As expected, the *QLR* statistic shows that the variables are jointly insignificant, while the *LR* statistic shows strong significance.

e. Using the robust standard errors does not significantly change any conclusions; in fact, most explanatory variables become slightly more significant than when we use the GLM standard errors. In this example, it is the adjustment by $\hat{\sigma} > 1$ that makes the most difference. Having fully robust standard errors has no additional effect.

f. We simply compute the turning point for the quadratic: $\hat{\beta}_{age}/(-2\hat{\beta}_{age^2})$ = 1143/(2*.00137) ≈ 41.72.

g. A double hurdle model -- which separates the initial decision to smoke at all from the decision of how much to smoke -- seems like a good idea. It is certainly worth investigating. One approach is to model $D(y|\mathbf{x}, y \geq 1)$ as a truncated Poisson distribution, and then to model $P(y = 0|\mathbf{x})$ as a logit or probit.

19.5. a. We just use iterated expectations:

$$E(y_{it}|\mathbf{x}_i) = E[E(y_{it}|\mathbf{x}_i, c_i)|\mathbf{x}_i] = E(c_i|\mathbf{x}_i)\exp(\mathbf{x}_{it}\boldsymbol{\beta})$$
$$= \exp(\alpha + \bar{\mathbf{x}}_i\boldsymbol{\gamma})\exp(\mathbf{x}_{it}\boldsymbol{\beta}) = \exp(\alpha + \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i\boldsymbol{\gamma}).$$

b. We are explicitly testing H_0: $\boldsymbol{\gamma} = \mathbf{0}$, but we are maintaining full independence of c_i and $\mathbf{x}_i$ under H_0. We have enough assumptions to derive $\text{Var}(\mathbf{y}_i|\mathbf{x}_i)$, the $T \times T$ conditional variance matrix of $\mathbf{y}_i$ given $\mathbf{x}_i$ under H_0. First,

$$\text{Var}(y_{it}|\mathbf{x}_i) = E[\text{Var}(y_{it}|\mathbf{x}_i, c_i)|\mathbf{x}_i] + \text{Var}[E(y_{it}|\mathbf{x}_i, c_i)|\mathbf{x}_i]$$
$$= E[c_i\exp(\mathbf{x}_{it}\boldsymbol{\beta})|\mathbf{x}_i] + \text{Var}[c_i\exp(\mathbf{x}_{it}\boldsymbol{\beta})|\mathbf{x}_i]$$
$$= \exp(\alpha + \mathbf{x}_{it}\boldsymbol{\beta}) + \tau^2[\exp(\mathbf{x}_{it}\boldsymbol{\beta})]^2,$$

where $\tau^2 \equiv \text{Var}(c_i)$ and we have used $E(c_i|\mathbf{x}_i) = \exp(\alpha)$ under H_0. A similar, general expression holds for conditional covariances:

$$\text{Cov}(y_{it}, y_{ir}|\mathbf{x}_i) = E[\text{Cov}(y_{it}, y_{ir}|\mathbf{x}_i, c_i)|\mathbf{x}_i]$$
$$+ \text{Cov}[E(y_{it}|\mathbf{x}_i, c_i), E(y_{ir}|\mathbf{x}_i, c_i)|\mathbf{x}_i]$$
$$= 0 + \text{Cov}[c_i\exp(\mathbf{x}_{it}\boldsymbol{\beta}), c_i\exp(\mathbf{x}_{ir}\boldsymbol{\beta})|\mathbf{x}_i]$$
$$= \tau^2\exp(\mathbf{x}_{it}\boldsymbol{\beta})\exp(\mathbf{x}_{ir}\boldsymbol{\beta}).$$

So, under H_0, $\text{Var}(\mathbf{y}_i|\mathbf{x}_i)$ depends on α, $\boldsymbol{\beta}$, and τ^2, all of which we can estimate. It is natural to use a score test of H_0: $\boldsymbol{\gamma} = \mathbf{0}$. First, obtain consistent estimators $\approx\!\!\!\!\alpha$, $\approx\!\!\!\!\boldsymbol{\beta}$ by, say, pooled Poisson QMLE. Let $\overset{\approx}{y}_{it} = \exp(\overset{\approx}{\alpha} + \mathbf{x}_{it}\overset{\approx}{\boldsymbol{\beta}})$ and $\overset{\approx}{u}_{it} = y_{it} - \overset{\approx}{y}_{it}$. A consistent estimator of τ^2 can be obtained from a simple pooled regression, through the origin, of

$$\overset{\approx}{u}{}^2_{it} - \overset{\approx}{y}_{it} \text{ on } [\exp(\mathbf{x}_{it}\overset{\approx}{\boldsymbol{\beta}})]^2, \; t = 1,\ldots,T; \; i = 1,\ldots,N.$$

Call this estimator $\tilde{\tau}^2$. This works because, under H_0, $E(u^2_{it}|\mathbf{x}_i) = E(u^2_{it}|\mathbf{x}_{it}) = \exp(\alpha + \mathbf{x}_{it}\boldsymbol{\beta}) + \tau^2[\exp(\mathbf{x}_{it}\boldsymbol{\beta})]^2$, where $u_{it} \equiv y_{it} - E(y_{it}|\mathbf{x}_{it})$. [We could also use the many covariance terms in estimating τ^2 because $\tau^2 =$

$\mathrm{E}\{[u_{it}^2/\exp(\mathbf{x}_{it}\boldsymbol{\beta})][u_{ir}^2/\exp(\mathbf{x}_{ir}\boldsymbol{\beta})]\}$, all $t \neq r$.

Next, we construct the $T \times T$ weighting matrix for observation i, as in Section 19.6.3; see also Problem 12.11. The matrix $\mathbf{W}_i(\tilde{\boldsymbol{\delta}}) = \mathbf{W}(\mathbf{x}_i,\tilde{\boldsymbol{\delta}})$ has diagonal elements $\tilde{\tilde{y}}_{it} + \tilde{\tau}^2[\exp(\mathbf{x}_{it}\tilde{\tilde{\boldsymbol{\beta}}})]^2$, $t = 1,\ldots,T$ and off-diagonal elements $\tilde{\tau}^2\exp(\mathbf{x}_{it}\tilde{\tilde{\boldsymbol{\beta}}})\exp(\mathbf{x}_{ir}\tilde{\tilde{\boldsymbol{\beta}}})$, $t \neq r$. Let $\tilde{\alpha}$, $\tilde{\boldsymbol{\beta}}$ be the solutions to

$$\min_{\alpha,\boldsymbol{\beta}} \ (1/2)\sum_{i=1}^{N}[\mathbf{y}_i - \mathbf{m}(\mathbf{x}_i,\alpha,\boldsymbol{\beta})]'[\mathbf{W}_i(\tilde{\boldsymbol{\delta}})]^{-1}[\mathbf{y}_i - \mathbf{m}(\mathbf{x}_i,\alpha,\boldsymbol{\beta})],$$

where $\mathbf{m}(\mathbf{x}_i,\alpha,\boldsymbol{\beta})$ has t^{th} element $\exp(\alpha + \mathbf{x}_{it}\boldsymbol{\beta})$. Since $\mathrm{Var}(\mathbf{y}_i|\mathbf{x}_i) = \mathbf{W}(\mathbf{x}_i,\boldsymbol{\delta})$, this is a MWNLS estimation problem with a correctly specified conditional variance matrix. Therefore, as shown in Problem 12.1, the conditional information matrix equality holds. To obtain the score test in the context of MWNLS, we need the score of the comditional mean function, with respect to all parameters, evaluated under H_0. Then, we can apply equation (12.69).

Let $\boldsymbol{\theta} \equiv (\alpha,\boldsymbol{\beta}',\boldsymbol{\gamma}')'$ denote the full vector of conditional mean parameters, where we want to test H_0: $\boldsymbol{\gamma} = \mathbf{0}$. The unrestricted conditional mean function, for each t, is

$$\mu_t(\mathbf{x}_i,\boldsymbol{\theta}) = \exp(\alpha + \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i\boldsymbol{\gamma}).$$

Taking the gradient and evaluating it under H_0 gives

$$\nabla_\theta\mu_t(\mathbf{x}_i,\tilde{\boldsymbol{\theta}}) = \exp(\tilde{\alpha} + \mathbf{x}_{it}\tilde{\boldsymbol{\beta}})[1,\mathbf{x}_{it},\bar{\mathbf{x}}_i],$$

which would be $1 \times (1 + 2K)$ without any redundancies in $\bar{\mathbf{x}}_i$. Usually, $\mathbf{x}_{it}$ would contain year dummies or other aggregate effects, and these would be dropped from $\bar{\mathbf{x}}_i$; we do not make that explicit here. Let $\nabla_\theta\boldsymbol{\mu}(\mathbf{x}_i,\tilde{\boldsymbol{\theta}})$ denote the $T \times (1 + 2K)$ matrix obtained from stacking the $\nabla_\theta\mu_t(\mathbf{x}_i,\tilde{\boldsymbol{\theta}})$ from $t = 1,\ldots,T$. Then the score function, evaluate at the null estimates $\tilde{\boldsymbol{\theta}} \equiv (\tilde{\alpha},\tilde{\boldsymbol{\beta}}',\tilde{\boldsymbol{\gamma}}')'$, is

$$\mathbf{s}_i(\tilde{\boldsymbol{\theta}}) = -\nabla_\theta\boldsymbol{\mu}(\mathbf{x}_i,\tilde{\boldsymbol{\theta}})'[\mathbf{W}_i(\tilde{\boldsymbol{\delta}})]^{-1}\tilde{\mathbf{u}}_i,$$

where $\tilde{\mathbf{u}}_i$ is the $T \times 1$ vector with elements $\tilde{u}_{it} \equiv y_{it} - \exp(\tilde{\alpha} + \mathbf{x}_{it}\tilde{\boldsymbol{\beta}})$. The estimated conditional Hessian, under H_0, is

$$\tilde{\mathbf{A}} = N^{-1}\sum_{i=1}^{N}\nabla_\theta\boldsymbol{\mu}(\mathbf{x}_i,\tilde{\boldsymbol{\theta}})'[\mathbf{W}_i(\tilde{\boldsymbol{\delta}})]^{-1}\nabla_\theta\boldsymbol{\mu}(\mathbf{x}_i,\tilde{\boldsymbol{\theta}}),$$

a $(1 + 2K) \times (1 + 2K)$ matrix. The score or LM statistic is therefore

$$LM = \left(\sum_{i=1}^{N}\nabla_\theta\boldsymbol{\mu}(\mathbf{x}_i,\tilde{\boldsymbol{\theta}})'[\mathbf{W}_i(\tilde{\boldsymbol{\delta}})]^{-1}\tilde{\mathbf{u}}_i\right)'\left(\sum_{i=1}^{N}\nabla_\theta\boldsymbol{\mu}(\mathbf{x}_i,\tilde{\boldsymbol{\theta}})'[\mathbf{W}_i(\tilde{\boldsymbol{\delta}})]^{-1}\nabla_\theta\boldsymbol{\mu}(\mathbf{x}_i,\tilde{\boldsymbol{\theta}})\right)^{-1} \cdot\left(\sum_{i=1}^{N}\nabla_\theta\boldsymbol{\mu}(\mathbf{x}_i,\tilde{\boldsymbol{\theta}})'[\mathbf{W}_i(\tilde{\boldsymbol{\delta}})]^{-1}\tilde{\mathbf{u}}_i\right).$$

Under H_0, and the full set of maintained assumptions, $LM \overset{a}{\sim} \chi^2_K$. If only $J < K$ elements of $\bar{\mathbf{x}}_i$ are included, then the degrees of freedom gets reduced to J.

In practice, we might want a robust form of the test that does not require $\mathrm{Var}(\mathbf{y}_i|\mathbf{x}_i) = \mathbf{W}(\mathbf{x}_i,\boldsymbol{\delta})$ under H_0, where $\mathbf{W}(\mathbf{x}_i,\boldsymbol{\delta})$ is the matrix described above. This variance matrix was derived under pretty restrictive assumptions. A fully robust form is given in equation (12.68), where $\mathbf{s}_i(\tilde{\boldsymbol{\theta}})$ and $\tilde{\mathbf{A}}$ are as given above, and $\tilde{\mathbf{B}} = N^{-1}\sum_{i=1}^{N}\mathbf{s}_i(\tilde{\boldsymbol{\theta}})\mathbf{s}_i(\tilde{\boldsymbol{\theta}})'$. Since the restrictions are written as $\boldsymbol{\gamma} = \mathbf{0}$, we take $\mathbf{c}(\boldsymbol{\theta}) = \boldsymbol{\gamma}$, and so $\tilde{\mathbf{C}} = [\mathbf{0}|\mathbf{I}_K]$, where the zero matrix is $K \times (1 + K)$.

c. If we assume (19.60), (19.61) and $c_i = a_i\exp(\alpha + \bar{\mathbf{x}}_i\boldsymbol{\gamma})$ where $a_i|\mathbf{x}_i \sim \mathrm{Gamma}(\delta,\delta)$, then things are even easier -- at least if we have software that estimates random effects Poisson models. Under these assumptions, we have

$y_{it}|\mathbf{x}_i,a_i \sim \mathrm{Poisson}[a_i\exp(\alpha + \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i\boldsymbol{\gamma})]$

y_{it}, y_{ir} are independent conditional on $(\mathbf{x}_i,a_i)$, $t \neq r$

$a_i|\mathbf{x}_i \sim \mathrm{Gamma}(\delta,\delta)$.

In other words, the full set of random effects Poisson assumptions holds, but where the mean function in the Poisson distribution is $a_i\exp(\alpha + \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i\boldsymbol{\gamma})$. In practice, we just add the (nonredundant elements of) $\bar{\mathbf{x}}_i$ in each time period, along with a constant and $\mathbf{x}_{it}$, and carry out a random effects Poisson analysis. We can test H_0: $\boldsymbol{\gamma} = \mathbf{0}$ using the LR, Wald, or score approaches. Any of these wouldbe asymptotically efficient. But none is robust because we have used a full distribution for $\mathbf{y}_i$ given $\mathbf{x}_i$.

19.7. a. First, for each t, the density of y_{it} given $(\mathbf{x}_i = \mathbf{x},\ c_i = c)$ is

$$f(y_t|\mathbf{x},c;\boldsymbol{\beta}_o) = \exp[-c\cdot m(\mathbf{x}_t,\boldsymbol{\beta}_o)][c\cdot m(\mathbf{x}_t,\boldsymbol{\beta}_o)]^{y_t}/y_t!,\quad y_t = 0,1,2,\ldots.$$

Multiplying these together gives the joint density of $(y_{i1},\ldots,y_{iT})$ given $(\mathbf{x}_i = \mathbf{x},\ c_i = c)$. Taking the log, plugging in the observed data for observation i, and dropping the factorial term gives

$$\sum_{t=1}^{T}\{-c_i m(\mathbf{x}_{it},\boldsymbol{\beta}) + y_{it}[\log(c_i) + \log(m(\mathbf{x}_{it},\boldsymbol{\beta}))]\}.$$

b. Taking the derivative of $\ell_i(c_i,\boldsymbol{\beta})$ with respect to c_i, setting the result to zero, and rerranging gives

$$(n_i/c_i) = \sum_{t=1}^{T} m(\mathbf{x}_{it},\boldsymbol{\beta}).$$

Letting $c_i(\boldsymbol{\beta})$ denote the solution as a function of $\boldsymbol{\beta}$, we have $c_i(\boldsymbol{\beta}) = n_i/M_i(\boldsymbol{\beta})$, where $M_i(\boldsymbol{\beta}) \equiv \sum_{t=1}^{T} m(\mathbf{x}_{it},\boldsymbol{\beta})$. The second order sufficient condition for a maximum is easily seen to hold.

c. Plugging the solution from part b into $\ell_i(c_i,\boldsymbol{\beta})$ gives

$$\begin{aligned}\ell_i[c_i(\boldsymbol{\beta}),\boldsymbol{\beta}] &= -[n_i/M_i(\boldsymbol{\beta})]M_i(\boldsymbol{\beta}) + \sum_{t=1}^{T} y_{it}\{\log[n_i/M_i(\boldsymbol{\beta})] + \log[m(\mathbf{x}_{it},\boldsymbol{\beta})]\\ &= -n_i + n_i\log(n_i) + \sum_{t=1}^{T} y_{it}\{\log[m(\mathbf{x}_{it},\boldsymbol{\beta})/M_i(\boldsymbol{\beta})]\\ &= \sum_{t=1}^{T} y_{it}\log[p_t(\mathbf{x}_i,\boldsymbol{\beta})] + (n_i - 1)\log(n_i),\end{aligned}$$

because $p_t(\mathbf{x}_i,\boldsymbol{\beta}) = m(\mathbf{x}_{it},\boldsymbol{\beta})/M_i(\boldsymbol{\beta})$ [see equation (19.66)].

d. From part c it follows that if we maximize $\sum_{i=1}^{N}\ell_i(c_i,\boldsymbol{\beta})$ with respect to $(c_1,\ldots,c_N)$ -- that is, we concentrate out these parameters -- we get exactly $\sum_{i=1}^{N}\ell_i[c_i(\boldsymbol{\beta}),\boldsymbol{\beta}]$. But, except for the term $\sum_{i=1}^{N}(n_i - 1)\log(n_i)$ -- which does not depend on $\boldsymbol{\beta}$ -- this is exactly the conditional log likelihood for the conditional multinomial distribution obtained in Section 19.6.4. Therefore, this is another case where treating the c_i as parameters to be estimated leads us to a $\sqrt{N}$-consistent, asymptotically normal estimator of $\boldsymbol{\beta}_o$.

19.9. I will use the following Stata output. I first converted the dependent variable to be in [0,1], rather than [0,100]. This is required to easily use the "glm" command in Stata.

```
. replace atndrte = atndrte/100
(680 real changes made)

. reg atndrte ACT priGPA frosh soph

      Source |       SS       df       MS              Number of obs =     680
-------------+------------------------------           F(  4,   675) =   72.92
       Model |  5.95396289     4  1.48849072           Prob > F      =  0.0000
    Residual |  13.7777696   675  .020411511           R-squared     =  0.3017
-------------+------------------------------           Adj R-squared =  0.2976
       Total |  19.7317325   679  .029059989           Root MSE      =  .14287

------------------------------------------------------------------------------
     atndrte |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ACT |  -.0169202    .001681   -10.07   0.000    -.0202207   -.0136196
      priGPA |   .1820163   .0112156    16.23   0.000     .1599947    .2040379
       frosh |   .0517097   .0173019     2.99   0.003     .0177377    .0856818
        soph |   .0110085    .014485     0.76   0.448    -.0174327    .0394496
       _cons |   .7087769   .0417257    16.99   0.000     .6268492    .7907046
------------------------------------------------------------------------------

. predict atndrteh
(option xb assumed; fitted values)

. sum atndrteh

    Variable |     Obs        Mean   Std. Dev.       Min        Max
-------------+-----------------------------------------------------
    atndrteh |     680    .8170956    .0936415   .4846666   1.086443

. count if atndrteh > 1
   12

. glm atndrte ACT priGPA frosh soph, family(binomial) sca(x2)
note: atndrte has non-integer values

Iteration 0:   log likelihood = -226.64509
Iteration 1:   log likelihood = -223.64983
Iteration 2:   log likelihood = -223.64937
Iteration 3:   log likelihood = -223.64937

Generalized linear models                          No. of obs      =       680
```

```
Optimization     : ML: Newton-Raphson                 Residual df     =        675
                                                      Scale param     =          1
Deviance         =  285.7371358                       (1/df) Deviance =   .4233143
Pearson          =  85.57283238                       (1/df) Pearson  =   .1267746

Variance function: V(u) = u*(1-u)                     [Bernoulli]
Link function    : g(u) = ln(u/(1-u))                 [Logit]
Standard errors  : OIM

Log likelihood   = -223.6493665                       AIC             =   .6724981
BIC              =  253.1266718

------------------------------------------------------------------------------
     atndrte |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ACT |  -.1113802   .0113217    -9.84   0.000    -.1335703   -.0891901
      priGPA |   1.244375   .0771321    16.13   0.000     1.093199    1.395552
       frosh |   .3899318    .113436     3.44   0.001     .1676013    .6122622
        soph |   .0928127   .0944066     0.98   0.326    -.0922209    .2778463
       _cons |   .7621699   .2859966     2.66   0.008      .201627    1.322713
------------------------------------------------------------------------------
(Standard errors scaled using square root of Pearson X2-based dispersion)

. di (.1268)^2
.01607824

. di exp(.7622 - .1114*30 + 1.244*3)/(1 + exp(.7622 - .1114*30 + 1.244*3))
.75991253

. di exp(.7622 - .1114*25 + 1.244*3)/(1 + exp(.7622 - .1114*25 + 1.244*3))
.84673249

. di .760 - .847
-.087

. predict atndh
(option mu assumed; predicted mean atndrte)

. sum atndh

    Variable |     Obs        Mean    Std. Dev.       Min        Max
-------------+-----------------------------------------------------
       atndh |     680    .8170956    .0965356   .3499525   .9697185

. corr atndrte atndh
(obs=680)

             |  atndrte    atndh
-------------+------------------
     atndrte |   1.0000
       atndh |   0.5725   1.0000
```

```
. di (.5725)^2
.32775625
```

a. The coefficient on *ACT* means that if the ACT score increases by 5 points -- more than a one standard deviation increase -- then the attendance rate is estimated to fall by about .017(5) = .085, or 8.5 percentage points. The coefficient on *priGPA* means that if prior GPA is one point higher, the attendance rate is predicted to be about .182 higher, or 18.2 percentage points. Naturally, these changes do not always make sense when starting at extreme values of *atndrte*. There are 12 fitted values greater than one; none less than zero.

b. The GLM standard errors are given in the output. Note that $\hat{\sigma} \approx .0161$. In other words, the usual MLE standard errors, obtained, say, from the expected Hessian of the quasi-log likelihood, are much too *large*. The standard errors that account for $\sigma^2 < 1$ are given by the GLM output. (If you omit the "sca(x2)" option in the "glm" command, you will get the usual MLE standard errors.)

c. Since the coefficient on *ACT* is negative, we know that an increase in ACT score, holding year and prior GPA fixed, actually reduces predicted attendance rate. The calculation shows that when *ACT* increases from 25 to 30, the estimated fall in *atndrte* is about .087, or about 8.7 percentage points. This is very similar to that found using the linear model.

d. The *R*-squared for the linear model is about .302. For the logistic functional form, I computed the squared correlation between $atndrte_i$ and $\hat{E}(atndrte_i|\mathbf{x}_i)$. This *R*-squared is about .328, and so the logistic functional form does fit better than the linear model. And, remember that the parameters in the logistic functional form are not chosen to maximize an *R*-squared.

19.11. To be added.

SOLUTIONS TO CHAPTER 20 PROBLEMS

20.1. To be added.

20.3. a. If all durations in the sample are censored, $d_i = 0$ for all i, and so the log-likelihood is $\sum_{i=1}^{N} \log[1 - F(t_i|\mathbf{x}_i;\boldsymbol{\theta})] = \sum_{i=1}^{N} \log[1 - F(c_i|\mathbf{x}_i;\boldsymbol{\theta})]$

b. For the Weibull case, $F(t|\mathbf{x}_i;\boldsymbol{\theta}) = 1 - \exp[-\exp(\mathbf{x}_i\boldsymbol{\beta})t^{\alpha}]$, and so the log-likelihood is $-\sum_{i=1}^{N} \exp(\mathbf{x}_i\boldsymbol{\beta})c_i^{\alpha}$.

c. Without covariates, the Weibull log-likelihood with complete censoring is $-\exp(\beta)\sum_{i=1}^{N} c_i^{\alpha}$. Since $c_i > 0$, we can choose any $\alpha > 0$ so that $\sum_{i=1}^{N} c_i^{\alpha} > 0$. But then, for any $\alpha > 0$, the log-likelihood is maximized by minimizing $\exp(\beta)$ across β. But as $\beta \to -\infty$, $\exp(\beta) \to 0$. So plugging any value α into the log-likelihood will lead to β getting more and more negative without bound. So no two real numbers for α and β maximize the log likelihood.

d. It is not possible to estimate duration models from flow data when all durations are right censored.

20.5. a. $P(t_i^* \leq t|\mathbf{x}_i,a_i,c_i,s_i = 1) = P(t_i^* \leq t|\mathbf{x}_i,t_i^* > b - a_i) = P(t_i^* \leq t, t_i^* > b - a_i|\mathbf{x}_i)/P(t_i^* > b - a_i|\mathbf{x}_i) = P(t_i^* \leq t|\mathbf{x}_i)/P(t_i^* > b - a_i|\mathbf{x}_i)$ (because $t < b - a_i$) $= [F(t|\mathbf{x}_i) - F(b - a_i|\mathbf{x}_i)]/[1 - F(b - a_i|\mathbf{x}_i)]$.

b. The derivative of the cdf in part a, with respect to t, is simply $f(t|\mathbf{x}_i)/[1 - F(b - a_i|\mathbf{x}_i)]$.

c. $P(t_i = c_i|\mathbf{x}_i,a_i,c_i,s_i = 1) = P(t_i^* \geq c_i|\mathbf{x}_i,t_i^* > b - a_i) = P(t_i^* \geq$

$c_i|\mathbf{x}_i)/\mathrm{P}(t_i^* \geq b - a_i|\mathbf{x}_i)$ (because $c_i > b - a_i$) $= [1 - F(c_i|\mathbf{x}_i)]/[1 - F(b - a_i|\mathbf{x}_i)]$.

20.7. a. We suppress the parameters in the densities. First, by (20.22) and $\mathrm{D}(a_i|c_i,\mathbf{x}_i) = \mathrm{D}(a_i|\mathbf{x}_i)$, the density of (a_i,t_i^*) given $(c_i,\mathbf{x}_i)$ does not depend on c_i and is given by $k(a|\mathbf{x}_i)f(t|\mathbf{x}_i)$ for $0 < a < b$ and $0 < t < \infty$. This is also the conditional density of (a_i,t_i) given $(c_i,\mathbf{x}_i)$ when $t < c_i$, that is, the observation is uncensored. For $t = c_i$, the density is $k(a|\mathbf{x}_i)[1 - F(c_i|\mathbf{x}_i)]$, by the usual right censoring argument. Now, the probability of observing the random draw $(a_i,c_i,\mathbf{x}_i,t_i)$, conditional on $\mathbf{x}_i$, is $\mathrm{P}(t_i^* \geq b - a_i,\mathbf{x}_i)$, which is exactly (20.32). From the standard result for densities for truncated distributions, the density of (a_i,t_i) given $(c_i,d_i,\mathbf{x}_i)$ and $s_i = 1$ is

$$k(a|\mathbf{x}_i)[f(t|\mathbf{x}_i)]^{d_i}[1 - F(c_i|\mathbf{x}_i)]^{(1 - d_i)}/\mathrm{P}(s_i = 1|\mathbf{x}_i),$$

for all combinations (a,t) such that $s_i = 1$. Putting in the parameters and taking the log gives (20.56).

b. We have the usual tradeoff between robustness and efficiency. Using the log likelihood (20.56) results in more efficient estimators provided we have the two densities correctly specified; (20.30) requires us to only specify $f(\cdot|\mathbf{x}_i)$.

20.9. To be added.